

Method of Displaying a Digital Image

TNS A1

This invention relates to a method of displaying a digital image.

I> A2

5 Static images are increasingly being stored in digital form, for example in computer memories. This is partly due to the increase in use of computer scanners and digital camera equipment. One of the main advantages of providing an image, e.g. a photograph, in digital form, is that it can be edited on a computer, displayed on a computer screen, or, with suitable equipment, displayed on a television screen.

10

5w> P3 Owing to the fact that such digital images (as with all static images) inherently hold a viewer's attention for a limited period of time, methods have been devised for producing a moving image from a static image. Such methods are commonly referred to as 'rostrum camera' techniques. A conventional rostrum camera is a film or television camera mounted vertically on a fixed or adjustable column, typically used for shooting graphics or animation – these techniques for producing moving images are of the type that can typically be obtained from such a camera. Essentially, these techniques involve different parts of the static image being displayed over time to provide an overall motion effect 'over' the image. For example, a perceivable panning motion from one part of the image to another might be employed. As a further example, a zooming effect might be used. The main purpose of using such rostrum camera techniques is to generate continued interest in an image by converting a static image into a moving image (or rather, a series of consecutive static images so arranged to represent motion). Rostrum camera techniques also have advantages in terms of displayable resolution. By moving over an image, detail of an image can be shown that otherwise cannot be shown on a low resolution display (without zooming-in).

Such conventional rostrum camera techniques are generally manual, in that they require a user to observe the displayed image, and manually to plot the path or order of the image parts to be displayed in a moving sequence. This can be time consuming, and requires some technical knowledge of camera equipment or digital image editing software. In order to provide a good-quality moving image, particularly for photographs of real-life scenes, some knowledge of photographic composition is also

required. This is clearly disadvantageous to an average person who wishes to add some degree of interest and variation to, say, a still photograph.

Basic automation of rostrum camera techniques are provided in a few digital image editing packages, such as 'Photo-Vista' from MGI Software Corp. Such a package provides a virtual moving viewpoint of a static digital image by moving from one side to the other, by moving in a wave like motion, or by displaying random parts (much like a screensaver slide-show). However, no account is taken of the image content, and no photographic composition (even at a basic level) is accounted for.

10

In many conventional techniques, computer memory is taken-up with displaying uninteresting parts of images and of those image parts that are considered significant or interesting, only small sections may be shown with a 'cutting-off' effect. Also, in terms of digital examples, large amounts of time and computational resources are used.

20

According to a first aspect of the present invention, there is provided a method of displaying a digital image, the method comprising: acquiring a set of image data representative of a displayable static image; performing an analysis of the image data using a processing means to identify characteristics of the image content; and generating, in the processing means, a set of video data for output to a display, the video data representing displayable motion over the static image and being generated in accordance with the image content characteristics.

25

The method provides an effective way of automatically producing a moving viewpoint over an otherwise static image, with the video data (representing the moving viewpoint) being dependant on characteristics of the image content. Thus, what is actually shown in the static image is taken into account when generating the moving viewpoint. This provides for a much more effective video sequence and allows the layman to produce highly effective results without any knowledge of photographic composition.

30 Computer memory is used more efficiently to display those parts of the image in which the person viewing is likely to be interested.

It will be appreciated that the video data generated can be of any digital video format, and may make use of some of the more structured or interactive video playback formats which are emerging in the field. The video data need not simply relate to a stream of frames which are played sequentially, but could also relate to video programmes where
5 a viewer can browse the programme in a non-sequential manner. This is possible in a format called Synchronised Media Integration Language (SMIL). This format allows identification of clips and other structure so that the viewer can jump around the video. Preferably, the step of performing an analysis of the image data comprises determining which of a number of predefined image characteristics are present in the image, and the
10 step of generating the video data comprises executing an algorithm associated with those characteristics identified, the algorithm defining a rule or rules for generating a moving viewpoint over the image for display.

The step of performing an analysis of the image may further comprise identifying a
15 predefined image class wherein, in that image class, sub-parts of the image have predefined characteristics, and establishing index frames based on a close-up view of each identified sub-part, the step of generating the video data comprising executing an algorithm for determining a display path from one index frame to the next. In the step of generating the video data, the algorithm may further determine one or more of (a) the
20 order of the index frames to be displayed, (b) the amount of time for which each index frame is displayed, and (c) the nature of the transition between each index frame.

The step of identifying the predefined image class having sub-parts with predefined characteristics may comprise identifying regions of interest and performing a feature
25 recognition operation. The step of performing feature recognition may identify human facial features, the step of establishing index frames thereafter comprising forming index frames based on a close-up view of the facial features. Having identified human facial features, the step of performing feature recognition may further comprise comparing the facial features with a database of pre-stored facial features such that the
30 step of forming index frames is performed only for those facial features already present in the database. The step of generating the video data to establish a display path can comprise determining the orientation of the facial features, and generating a display path which follows the general gaze direction which the facial features exhibit.

2010 RELEASE UNDER E.O. 14176

As an alternative to the index-frame method above, the step of performing an analysis of the image may further comprise identifying a predefined image class wherein, in that image class, there are one or more dominant edges, lines or curves, the step of
5 generating the video data comprising executing an algorithm for determining a display path following the one or more dominant edges, lines or curves.

Further, the step of performing an analysis of the image data may further comprise: (a) identifying a predefined image class wherein, in that image class, there are both (i)
10 image sub-parts having predefined characteristics, and (ii) dominant edges, lines or curves; and (b) establishing index frames based on a close-up view of each identified image sub-part in (i), the step of generating the video data comprising executing an algorithm for determining a display path moving between each index frame and following the dominant edges, lines or curves.
15

In the step of generating the video data, the algorithm may define rules having a first level and at least one sub-level, the rules in the first level relating to identification of a predefined image class, and the rules in the at least one sub-level relating to options for generating the moving viewpoint for the image class identified. The method may
20 further comprise prompting the user manually to select an option in a sub-level.

The step of generating the video data may comprise generating video data for a plurality of video sub-clips, each sub-clip representing displayable motion over a different part of the static image, and wherein the method further comprises an editing
25 step for linking the sub-clips to form a second set of video data.

The above method finds particular application where the image data is representative of a displayable photograph.

30 According to a second aspect of the present invention, there is provided a computer program stored on a computer-readable medium, the computer program comprising computer-readable instructions for causing the computer to execute the steps of: acquiring a set of image data representative of a displayable static image; performing

an analysis of the image data using a processing means to identify characteristics of the image content; and generating, in the processing means, a set of video data for output to a display, the video data representing displayable motion over the static image and being generated in accordance with the image content characteristics.

5

According to a third aspect of the present invention, there is provided a computer system comprising a processing means, a data port and a video port, the processor being arranged to receive image data representative of a displayable static image from the data port, and wherein the processing means is further arranged to access and to 10 perform an analysis of the image data to identify characteristics of the image content, and to generate a set of video data representing displayable motion over the static image according to the image content characteristics, the processor being arranged to output the video data to the video port for display.

SEARCHED
INDEXED
SERIALIZED
FILED

A5

15 The invention will now be described, by way of example, with reference to the accompanying drawings in which:

Figure 1 is a block diagram representing a computer system for running a computer program according to the invention;

20

Figure 2 is a block diagram representation of a set of rules within the computer program for generating video data from a static image;

Figures 3a to 3d represent the stages of forming a video programme from a static 25 photograph depicting a group of people;

Figure 4 represents a photograph showing a landscape scene; and

Figure 5 represents a photograph showing a landscape scene with a subject being in the 30 foreground.

I >> A b

A first embodiment of the invention is shown in Figure 1. Referring to Figure 1, a computer system 1 comprises a processor 3 connected to a memory 5. The processor 3

receives data via an input port 7 and outputs data to a display 9, such as a monitor. In this case, the input port 7 receives image data from an image source, for example a scanner or a digital camera. The image data received represents a static image, here a photograph, and is stored in the memory 5. Also stored in the memory 5 is an application program 11 for generating a set of "video data" from the image data. The video data generated by the application program 11 provides a displayable "video programme" representing a displayable moving viewpoint over the static image. This video programme can thereafter be viewed on the display 9. It will be appreciated that such a video programme will actually be composed of a plurality of separate static images so sequenced to give the motion effect. For ease of explanation the terms video data and video programme will be referred to.

Once the acquired image data is stored in the memory 5, the application program 11 is launched, and the location of the file holding the image data specified to the program. The program 11, operating under the control of the processor 3, then performs an analysis of the image data such that particular characteristics of the image content (i.e. that content which can be viewed in two dimensions on a display) are identified. The application then proceeds to generate the abovementioned video data based on the identified image content characteristics. To facilitate this, the application program 11 includes an algorithm comprising a set of rules for determining how the video programme should display a moving viewpoint based on the different characteristics identified.

In the present embodiment, the analysis stage of the application program 11 initially identifies which one of a plurality of predefined image classes the image data stored actually represents. Given that the intended image data relates to photographs, three image classes are provided in the program 11, namely (a) a face/group traversal class, (b) a landscape traversal class, and (c) a combined face/landscape traversal class.

In (a), i.e. the face/group traversal class, the analysis stage of the application program 11 uses standard image analysis techniques to identify so-called 'regions of interest'. Such techniques effectively involve processing the image data to decompose and find 'interesting' points within the image. As an example, such a technique might include

EPO/EPO/EP2002/00000000000000000000000000000000

identifying and segmenting regions having substantially uniform texture and colour. A further known method is to perform face detection whereby human facial features are identified. This latter method is used in the application program 11 to identify facial features within the image data belonging to one or more persons captured in a photograph. Assuming one or more faces are identified, the program 11 establishes 'index frames' of those faces based on a zoomed-in view of the face.

In the next phase of the application program 11, i.e. the video data generation stage, a display path is established, which effectively represents the moving viewpoint over or between the one or more index frames that the video data will produce when displayed. As explained briefly above, the application program 11 comprises an algorithm, or set of rules, many of which are based on photographic composition techniques for producing appropriate and interesting effects according to the image content. Thus, for example, where one face is identified in an image, the video data may represent an overview shot, showing the entire image, followed by a zooming-in effect to the face. Where one or more faces are identified, the overview may be followed by a panning motion between each index frame (hence group traversal class), if necessary, incorporating zooming-in and zooming-out effects to take account of some faces being closer than others. Of course, various other rules may be applied to obtain suitable motion effects. These will be discussed in more detail below.

In (b), i.e. the landscape traversal class, the analysis stage of the application program 11 identifies dominant edges, curves or lines in an image. For example, in a landscape scene having hills and mountains in the distance, the program 11 identifies the dominant edges, curves or lines (i.e. as being significantly more dominant than any particular region of interest, if any are present) and so knows to operate in this class in the video data generation stage. In this stage, the video data generated will follow the general path of the dominant edges, curves or lines in the image (hence landscape traversal class). As with the group traversal class, other rules may be applied to provide more sophisticated motion effects.

In (c), i.e. the combined face/landscape traversal class, the analysis stage identifies image data representing an image having both (i) one or more regions of interest and

(ii) a dominant edge, line or curve. In effect, it combines the analysis stages of (a) and (b) above. In this case, the video data generation stage will plot a path for showing the identified regions of interest whilst also following the general path of the edge, line or curve. This is a slightly more complex task, but all the same, provides for interesting
5 moving image effects.

Figure 2 shows a block diagram representing the general operation of the application program 11. As described above, in the initial analysis stage, the program 11 establishes into which of the three classes the image data falls. Having performed this,
10 the video generation stage will proceed to operate according to the predefined algorithm, or set of rules, belonging to the class identified (and represented as blocks 13, 15 and 17). Sub-rules are provided which are used to determine how the video data will be generated in that class. In Figure 2, sub-rules are shown and described with respect to the group traversal class only, although it will be appreciated that suitable
15 sub-rules may be formulated for the landscape traversal and combined face/landscape transversal classes.

An example image is shown in Figure 3a, the image representing a photograph 27 of a group of people. In the analysis stage of the application program 11, regions of interest
20 are identified in the image firstly by identifying areas having consistent colour and texture, and then by identifying the facial features. The program 11 then makes the decision that this image must belong to the group traversal class since it contains a number of distinct regions of interest (although it should be noted that even a single face or other region of interest would fall into this class also). Having decided on the
25 appropriate image class, the program 11 proceeds to generate a video programme based on the sub-rules shown in Figure 2.

Referring back to Figure 2, the first sub-rule 19 in the group traversal class relates to index frame selection. This selection is used to derive a set of still frames, or index
30 frames. Having identified that there are a number of faces in the analysis stage, the obvious choice is to use a close-up version of each face as the index frames. This is illustrated in Figure 3b, where index frames are represented as 27a - 27d. However, it should be appreciated that the choice of index frames may encompass more

sophisticated selection rules based on photographic composition techniques. These would comprise extra sub-levels (not shown) depending from the index frame selection rule in Figure 2. Such selection techniques are described in the applicant's co-pending International Patent Application No. GB01/05683 entitled "Automated Cropping of Electronic Images", filed on 20 December 2001, the contents of which are incorporated herein by reference. Image content features that are relevant to composition include face identification (used in this example), person identification (i.e. comparing an identified face with a database of pre-stored faces and using only recognised faces as index frames), gaze direction (faces appearing to look at each other providing interesting subjects), avoiding flat, boring and strongly contrasting areas, avoiding light areas adjacent the edges of the image etc.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

The second sub-rule 21 relates to the order in which the index frames will be shown. In the example shown, the rule 21 simply stipulates that a panning effect will take place from left to right, passing over each index frame in turn. This is shown in Figures 3c and 3d. Since the face on the extreme right hand side of the image is nearer, and so bigger, a zooming effect is used.

Further examples include the order following other rules of composition, for example, starting with an overview shot encompassing several regions of interest, and then showing detailed views of each region of interest. The overview shot sets the location and gives atmospheric effect. Another approach might withhold the location information from the user or audience to give some dramatic effect, the individual index frames being shown first before revealing the overview shot, i.e. by zooming-out. The ordering of index frames could follow the geometry of the scene. If the scene had shown a group of people in rows, e.g. a school photograph, the order could start from the top left of the scene, move rightwards, then down, and then leftwards to end at the bottom left person. Detecting the gaze of people (mentioned above) in the scene can be used to plot a suitable path over the image. Some degree of suspense can be generated since the viewer will wonder what the person is looking at. The part of the image the people are actually looking at can be kept out of view until the end.

The main effects which generate reaction from a viewer, and so are taken into account in this sub-rule, include panning for indicating spatial relationships between two or

more index frames (cutting effects do provide the same sense of continuity). Tilting upwards creates feelings of rising interest and emotion, expectancy, hope or anticipation. Tilting downwards is allied to lowering of interest and emotion, disappointment, sadness or critical inspection. Zooming-in or zooming-out indicates 5 the viewer's behaviour towards the subject. Rhythmic or jerky zooms should be avoided, so any zooming effect should be used with care.

The third sub-rule 23 determines the time period for which each index frame 27a - 27d is displayed. As the video programme moves between the index frames 27a - 27d, the 10 camera will linger at each index frame prior to moving to the next, in this case rightwards, index frame. The longer the time period, the more time the viewer is able to look at the detail in the index frame 27a - 27d. Ultimately, the time spent viewing each index frame 27a - 27d will depend on the number of index frames and the total time allotted to the video, say 20 seconds. For example, each index frame 27a - 27d 15 could be viewed static for approximately 2 seconds before the viewpoint moves to the next index frame. Of course, it is possible to use repeated views of the index frames 27a - 27d.

The fourth sub-rule 25 determines the nature of the movement between the index 20 frames 27a - 27d. This movement can vary according to the route (determined by the order) and the rate of movement (to some extent, determined in the previous stage). In Figure 3c, the motion between the index frames 27a - 27d is formed by straight lines between the centre of each index frame. In Figure 3d, a smoother, less jerky path is formed. The path could follow some structure in the image. Another approach is to 25 use a single line as a path through the people, with the size of the frame sufficient to capture all of the people of interest whilst panning the scene. In this sense, displayable resolution is compromised to some extent for smooth camera motion. A simple way of performing this would be to fit a line through a series of points (known as 'least squares fitting'). So, if a group of people were shown, a point could be made on each person 30 (say, their left eye). Next, a frame size is used which is sufficient to include each of the people (plus some margin space) when the camera viewpoint is at the point on the line closest to the fitted line. Complexity occurs when there are multiple rows of people of

DRAFT DOCUMENT

different sizes. One might try to fit several lines. If the variation in size is too great, then the frame size will become too large and the effect of seeing detail will be lost.

Other arrangements for panning between people or faces can be arranged. For
 5 example, if the application program 11 recognises people around an object, such as a table (e.g. with people along both longitudinal sides of the table, and at the opposite end to the viewpoint, then fitting horizontal lines to the image will result in a video programme that zig-zags across the table. However, by fitting three sides of a trapezium to points on the people, the camera viewpoint will move around the table,
 10 producing a much more desirable result.

Generally speaking, the speed of transition between index frames may depend on the amount of interest in the regions between index frames. Panning slowly across uninteresting areas can prove annoying to the viewer. The panning should be smooth,
 15 with erratic or hesitant panning avoided. A fast pan moving rapidly from one index frame to the next causes the intermediate region to appear blurred. As the viewer's attention is moved rapidly to the next index frame, it gives it (the next frame) a transitory importance.

20 A further example image is shown in Figure 4, the image representing a photograph 29 of a mountain range. In the analysis stage of the application program 11, no identifiable regions of interest are identified in the image, however, a dominant line is detected in the form of the mountain extremes contrasting with the background. The program 11 then makes the decision that this image must belong to the landscape
 25 traversal class since it contains no distinct regions of interest but there is a dominant line. Having decided on the appropriate image class, the program 11 proceeds to generate a video programme based on sub-rules (not shown in Figure 2). At the most basic level, the video programme shows a panning motion following the dominant edge (i.e. a zig-zag motion). Sub-rules may determine the direction and speed of motion.

30

Iws A7 A third example image is shown in Figure 5, the image representing a photograph 31 of a mountain range with a person positioned in the foreground. In the analysis stage of the application program 11, one identifiable region of interest is identified in the image,

namely the face. However, a dominant line is also detected in the form of the mountain extremes contrasting with the background (as in the previous example). The program 11 makes the decision that this image belongs to the combined face/landscape traversal class. Having decided on the appropriate image class, the program 11 proceeds to 5 generate a video programme based on suitable sub-rules (not shown in Figure 2). Again, the video programme may show a panning motion following the dominant edge, which zooms-out to show the index frame of the face. Alternatively, the reverse motion can be applied.

- 10 In the above embodiment, the application program 11 produces a video programme automatically from the image data. However, in second and third embodiments, some interaction is provided in the application program 11. This interaction allows a user to analyse and assist the specification of the moving viewpoint.
- 15 In the second embodiment (first interactive example), the application program 11 is used (in its automatic mode) to produce an initial video programme. This video programme is then used for analysis purposes and subsequent interaction. For example, index frames can be added or removed. The order of traversal of the index frames can be changed. The zoom around the index frames can also be altered. The application 20 program 11 provides an interface whereby these interactive options can be made. For example, a button can be used to signal that the viewable frame should zoom-in or zoom-out on a particular position. The rate of movement can also be altered. A button could also be used to speed up or to slow down the movement between index frames.
- 25 In the third embodiment (second interactive example), having decided on the image class in the analysis stage, instead of proceeding to generate the video data, the user is prompted to make certain decisions with respect to the sub-rules shown in Figure 2. For example, the user may indicate a point of interest in the image, the system thereafter using segmentation to deduce the region of interest and the required frame.
- 30 The automated part of the application program 11 can then calculate a suitable path for a viewpoint over the image using the or each region of interest. The order can also be specified. As a specific example, in the group traversal class 13, the user may be prompted to select or deselect some of the identified index frames 27a - 27d, to specify

the order of the index frames, to vary the time for which each index frame is specified, or to specify the nature of transition between each index frame. Although this may reduce the composition effects pre-specified in the program 11, a user is able to make some choices to suit individual requirements.

5

- In a fourth embodiment, a number of separate video programmes are generated within a single image for subsequent editing to form an overall video programme. For example, a moving viewpoint could be automatically generated for two or more different parts of the image. In the editing stage, the transition between the two programmes can be
10 specified manually.

HOBSONS LIBRARY

- In a fifth embodiment, the application program 11 is specifically arranged to analyse image data relating to photographs, and uses motion (in this example, panning) to overcome some aspect ratio mismatches between the photograph and the visual display.
15 This is useful since photographs tend to have different aspect ratios to conventional television screens. This means that if the whole photograph is displayed, there will be unused areas of the screen. This is a particular problem for portrait mode photographs. In this embodiment, the image data is analysed to detect the salient parts of the image. These are then used to form an overview frame. This overview frame forms the first
20 and last frame of the video sequence. The application program 11 generates a video sequence which is a panning effect about the image in such a way that it ends with the salient image. The panning action ensures that the whole image can be viewed.

- In all the above embodiments, use can be made of known visual techniques, such as
25 fades, cuts, and dissolves, to join separate video clips together.

- It should be appreciated that there can be provided several implementation strategies for the computer system 1. The first case is where the image data is acquired and analysed in real-time and the video data generated on demand. The second case is
30 where the viewpoint path, scale and frame rate information is stored after the analysis of the image is performed. In a third example, the whole of the image data can be sub-sampled to obtain a number of resolutions forming, effectively, a pyramid of images, as well as storing all viewpoint information. Since sub-sampling is computationally

expensive, caching them in advance is advantageous. The fourth case simply generates a complete video stream that can be played back sequentially.

Such a method and system for displaying digital images enables interesting and
5 effective video sequences to be generated from otherwise static images. Less computer
memory is occupied with image data which represents those areas which are of little of
no interest to the viewer.

2005 RELEASE UNDER E.O. 14176